Exact and Monte Carlo Calculations of Integrated Likelihoods for the Latent Class Model

Christophe Biernacki, Gilles Celeux, Gérard Govaert

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Exact and Monte Carlo Calculations of Integrated Likelihoods for the Latent Class Model

Christophe Biernacki — Gilles Celeux — Gérard Govaert

N° 6609

Août 2008

Thème COG

ISSN 0249-6399 ISRN INRIA/RR--6609--FR+ENG

Exact and Monte Carlo Calculations of Integrated Likelihoods for the Latent Class Model

Christophe Biernacki*, Gilles Celeux† , Gérard Govaert‡

Thème COG — Systèmes cognitifs
Équipes-Projets SELECT

Rapport de recherche n° 6609 — Août 2008 — 25 pages

Abstract: The latent class model or multivariate multinomial mixture is a powerful approach for clustering categorical data. This model uses a conditional independence assumption given the latent class to which an object is belonging to represent heterogeneous populations. . In this paper, we exploit the fact that a fully Bayesian analysis with Jeffreys non informative prior distributions does not involve technical difficulty to propose an exact expression of the integrated *complete-data* likelihood, which is known as being a meaningful model selection criterion in a clustering perspective. Similarly, a Monte Carlo approximation of the integrated *observed-data* likelihood can be obtained in two steps: An exact integration over the parameters is followed by an approximation of the sum over all possible partitions through either a frequentist or a Bayesian importance sampling strategy. Then, the exact and the approximate criteria experimentally compete respectively with their standard asymptotic BIC approximations for choosing the number of mixture components. Numerical experiments on simulated data and a biological example highlight that asymptotic criteria are usually dramatically more conservative than the non asymptotic presented criteria, not only for moderate sample sizes as expected but also for quite large sample sizes. It appears that asymptotic standard criteria could often fail to select some interesting structures present in the data. It is also the opportunity to highlight the deep purpose difference between the integrated *complete-data* and the *observed-data* likelihoods: The integrated *complete-data* likelihood is focussing on a cluster analysis view and favors well separated clusters, implying some robustness against model misspecification, while the integrated *observed-data* likelihood is focussing on a density estimation view and is expected to provide a consistent estimation of the distribution of the data.

Key-words: Categorical data, Bayesian model selection, Jeffreys conjugate prior, importance sampling, EM algorithm, Gibbs sampler

* CNRS & Université de Lille 1, Villeneuve d'Ascq
† Inria Saclay Île-de-France
‡ Université de Technologie de Compiègne

Centre de recherche INRIA Sophia Antipolis – Méditerranée
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex
Téléphone : +33 4 92 38 77 77 —Télécopie : +33 4 92 38 77 65

Calcul exact et par approximation de Monte-Carlo de vraisemblances intégrées pour le modèle des classes latentes

Résumé : Le modèle des classes latentes ou le modèle de mélange de lois multinomiales multivariées est un outil puissant pour la classification de données qualitatives. Ce modèle utilise pour représenter des populations hétérogènes une hypothèse d'indépendance conditionnelle sachant la classe d'un individu. Dans ce rapport, nous tirons parti du fait que, dans un cadre bayésien, la loi non informative de Jeffreys est bien définie pour en déduire l'expression exacte de la vraisemblance complétée intégrée qui constitue un critère de classification efficace. On en tire une approximation non asymptotique de la vraisemblance intégré observée. Cette approximation se fait en exprimant cette quantité comme somme sur toutes les partitions possibles de la vraisemblance intégée complétée. Il est alors possible d'en fournir une approximation de Monte-Carlo. Des expérimentations sur des données simulées et réelles permettent de calculer ces crirtères non asymptotiques avec leurs pendants asymptotiques dont le critère BIC. Ces expérimentations illustrent le gain important de notre approche pour des tailles d'échantillons faibles. Cette étude nous donne l'occasion de marquer les différences d'objectif et de comportement des critères de type vraisemblance intégré complétée qui favorisent des modèles donnant lieu à des classifications stables et pertinentes des données avec des critères de vraisemblance intégrée observée qui recherchent des modèles présentant un bon ajustement aux données sans se préoccuper de classification.

Mots-clés : Variables qualitatives, sélection bayésienne de modèles, loi conjuguée non informative de Jeffreys, échantillonnage préférentiel, algorithme EM, échantillonnage de Gibbs

1 Introduction

The standard model for clustering observations described through categorical variables is the so-called latent class model (see for instance Goodman 1974). This model is assuming that the observations arose from a mixture of multivariate distributions and that knowing the clusters they are conditionally independent. It has been proved to be successful in many practical situations (see for instance Aitkin et al. 1981).

In this paper, we consider the problem of choosing a relevant latent class model. In the Gaussian mixture context, the BIC criterion (Schwarz 1978) appears to give a reasonable answer to the important problem of choosing the number of mixture components (see for instance Fraley and Raftery 2002). However, some previous works dealing with the latent class model (see for instance Nadif and Govaert 1998) for the binary case suggest that BIC needs particular large sample size to reach its expected asymptotic behavior in practical situations. And, any criterion related to the asymptotic BIC approximation may suffer this limitation. In this paper, we take profit from the possibility to avoid asymptotic approximation of integrated likelihoods to propose alternative non asymptotic criteria.

Actually, a conjugate Jeffreys non informative prior distribution for the latent class model parameters is available and integrating the complete-data likelihood leads to a closed form formula, contrary to what happens for Gaussian mixture models. Thus, the integrated *complete-data* likelihood proposed in Biernacki et al. (2000) as a Bayesian *clustering* criterion can be exactly and easily computed without needing any BIC approximation. Moreover, the *observed-data* likelihood (see for instance Robert 2001) can be non asymptotically approximated through two steps: An exact integration of the complete data distribution over the parameters is followed by an approximation of the sum over all possible partitions to get the marginal distribution of the observed data. This approximation involves either a frequentist or a Bayesian importance sampling strategy. The Bayesian instrumental distribution is derived in a natural way using the fact that Bayesian inference is efficiently implemented through a Gibbs sampler thanks to conjugate properties.

The aim of this paper is to present those non asymptotic Bayesian (latent class) model selection criteria. It is organised as follows. In Section 2, the standard latent class model is described; furthermore maximum likelihood (ml) and non informative Bayesian inferences are briefly sketched. The exact integrated *complete-data* likelihood and the approximate integrated *observed-data* likelihood are respectively described in Section 3 and Section 4. Numerical experiments on both simulated and real data sets for selecting a relevant number of mixture components are presented in Section 5. A discussion section ends the paper and gives some possible extensions of this work.

2 The latent class model

The model Observations to be classified are described with d discrete variables. Each variable j has m_j response levels. Data are $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ where $\mathbf{x}_i = (x_i^{jh}; j = 1, \ldots, d; h = 1, \ldots, m_j)$ with $x_i^{jh} = 1$ if i has response level h for variable j and $x_i^{jh} = 0$ otherwise. In the standard latent class model, data are

supposed to arise independently from a mixture of g multivariate multinomial distributions with probability distribution function (pdf)

$$\mathrm{p}(\mathbf{x}_i;\boldsymbol{\theta}) = \sum_{k=1}^{g} \pi_k \mathrm{p}(\mathbf{x}_i;\boldsymbol{\alpha}_k) \quad \text{with} \quad \mathrm{p}(\mathbf{x}_i;\boldsymbol{\alpha}_k) = \prod_{j=1}^{d}\prod_{h=1}^{m_j} (\alpha_k^{jh})^{x_i^{jh}}, \qquad (1)$$

where $\boldsymbol{\theta} = (\boldsymbol{\pi}, \boldsymbol{\alpha})$ is denoting the vector parameter of the latent class model to be estimated, with $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_g)$ the vector of mixing proportions of the g latent clusters, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_g)$ and $\boldsymbol{\alpha}_k = (\alpha_k^{jh}; j = 1, \ldots, d; h = 1, \ldots, m_j)$, α_k^{jh} denoting the probability that variable j has level h if object i is in cluster k. The latent class model is assuming that the variables are *conditionally independent* knowing the latent clusters.

Analysing multivariate categorical data is made difficult because of the curse of dimensionality. The standard latent class model which requires $(g-1) + g * \sum_j (m_j - 1)$ parameters to be estimated is an answer to the dimensionality problem. It is much more parsimonious than the saturated loglinear model which requires $\prod_j m_j$ parameters. For instance, with $g = 5$, $d = 10$, $m_j = 4$ for all variables, the latent class model is characterised with 154 parameters whereas the saturated loglinear model requires about 10^6 parameters. Moreover, the latent class model can appear to produce a better fit than unsaturated loglinear models while demanding less parameters.

Maximum likelihood inference Since the latent class structure is a mixture model, the EM algorithm (Dempster et al. 1977, McLachlan and Krishnan 1997) is a privileged tool to derive the ml estimates of this model parameters (see McLachlan and Peel 2000). The observed-data log-likelihood of the model is

$$L(\boldsymbol{\theta};\mathbf{x}) = \sum_{i=1}^{n} \log \left(\sum_{k=1}^{g} \pi_k \prod_{j=1}^{d}\prod_{h=1}^{m_j} (\alpha_k^{jh})^{x_i^{jh}} \right). \qquad (2)$$

Noting the unknown indicator vectors of the g clusters by $\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)$ with $\mathbf{z}_i = (z_{i1}, \ldots, z_{ig})$ where $z_{ik} = 1$ if $\mathbf{x}_i$ arose from cluster k, $z_{ik} = 0$ otherwise, the complete-data log-likelihood is

$$L(\boldsymbol{\theta};\mathbf{x},\mathbf{z}) = \sum_{i=1}^{n}\sum_{k=1}^{g} z_{ik} \log \left(\pi_k \prod_{j=1}^{d}\prod_{h=1}^{m_j} (\alpha_k^{jh})^{x_i^{jh}} \right). \qquad (3)$$

From this complete-data log-likelihood, the equations of the EM algorithm are easily derived and this algorithm is as follows from an initial position $\boldsymbol{\theta}^0 = (\boldsymbol{\pi}^0, \boldsymbol{\alpha}^0)$.

- E step: Calculation of the conditional probability $t_{ik}(\boldsymbol{\theta}^r)$ that $\mathbf{x}_i$ arose from cluster k $(i = 1, \ldots, n; k = 1, \ldots, g)$

$$t_{ik}(\boldsymbol{\theta}^r) = \frac{\pi_k^r \mathrm{p}(\mathbf{x}_i;\boldsymbol{\alpha}_k^r)}{\sum_{\ell=1}^{g} \pi_\ell^r \mathrm{p}(\mathbf{x}_i;\boldsymbol{\alpha}_\ell^r)}. \qquad (4)$$

- M step: Updating of the mixture parameter estimates,

$$\pi_k^{r+1} = \frac{\sum_i t_{ik}(\boldsymbol{\theta}^r)}{n} \quad \text{and} \quad (\alpha_k^{jh})^{r+1} = \frac{\sum_{i=1}^{n} t_{ik}(\boldsymbol{\theta}^r) x_i^{jh}}{\sum_{i=1}^{n} t_{ik}(\boldsymbol{\theta}^r)}. \qquad (5)$$

Bayesian inference Since the Jeffreys non informative prior distribution for a multinomial distribution $\mathcal{M}_g(\mathrm{p}_1, \ldots, \mathrm{p}_g)$ is a conjugate Dirichlet distribution $\mathcal{D}_g(\frac{1}{2}, \ldots, \frac{1}{2})$, a fully non informative Bayesian analysis is possible for latent class models contrary to the Gaussian mixture model situation (see for instance Marin et al. 2005). Thus, using the prior distribution $\mathcal{D}_g(\frac{1}{2}, \ldots, \frac{1}{2})$ for the mixing weights, and noting $n_k = \#\{i : z_{ik} = 1\}$, the full conditional distribution of $\boldsymbol{\pi}$ is given by

$$\mathrm{p}(\boldsymbol{\pi}|\mathbf{z}) = \mathcal{D}_g(\tfrac{1}{2} + n_1, \ldots, \tfrac{1}{2} + n_g). \tag{6}$$

In a similar way, using the prior distribution $\mathcal{D}_{m_j}(\frac{1}{2}, \ldots, \frac{1}{2})$ for $\boldsymbol{\alpha}_k^j = (\alpha_k^{j1}, \ldots, \alpha_k^{jm_j})$, with $k = 1, \ldots, g$ and $j = 1, \ldots, d$, the full conditional distribution for $\boldsymbol{\alpha}_k^j$ is, noting $n_k^{jh} = \#\{i : z_{ik} = 1, x_i^{jh} = 1\}$,

$$\mathrm{p}(\boldsymbol{\alpha}_k^j|\mathbf{x}, \mathbf{z}) = \mathcal{D}_{m_j}(\tfrac{1}{2} + n_k^{j1}, \ldots, \tfrac{1}{2} + n_k^{jm_j}). \tag{7}$$

Finally, since the conditional probabilities of the indicator vectors $\mathbf{z}_i$ are given, for $i = 1, \ldots, n$, by

$$\mathrm{p}(\mathbf{z}_i|\mathbf{x}_i, \boldsymbol{\theta}) = \mathcal{M}_g(t_{i1}(\boldsymbol{\theta}), \ldots, t_{ig}(\boldsymbol{\theta})), \tag{8}$$

the Gibbs sampling implementation of the fully non informative Bayesian inference is straightforwardly deduced from those formulas and is not detailed further here. In addition, since $\mathbf{z}$ is discrete and finite, the convergence of the chain on $\boldsymbol{\theta}$ towards the stationary distribution $\mathrm{p}(\boldsymbol{\theta}|\mathbf{x})$ is geometric (see Robert 2001, for instance).

Because the prior distribution is symmetric in the components of the mixture, the posterior distribution is invariant under a permutation of the component labels (see for instance McLachlan and Peel 2000, Chap. 4). This lack of identifiability of $\boldsymbol{\theta}$ corresponds to the so-called *label switching* problem. In order to deal with this problem, some authors as Stephens (2000) or Celeux et al. (2000) apply a clustering-like method to possibly change the component labels of the simulated values for $\boldsymbol{\theta}$. In the same spirit, an alternative strategy making use of the ml estimate $\hat{\boldsymbol{\theta}}$ will be used in this paper: For each simulated $\boldsymbol{\theta}$ the chosen label permutation is minimising the Kullback-Leibler divergence

$$\mathrm{KL}(\mathrm{p}(\mathbf{z}|\hat{\boldsymbol{\theta}}; \mathbf{x}), \mathrm{p}(\mathbf{z}|\boldsymbol{\theta}; \mathbf{x})) = \sum_{i,k} t_{ik}(\hat{\boldsymbol{\theta}}) \ln t_{ik}(\hat{\boldsymbol{\theta}}) - \sum_{i,k} t_{ik}(\hat{\boldsymbol{\theta}}) \ln t_{ik}(\boldsymbol{\theta}). \tag{9}$$

between the conditional distributions $\mathrm{p}(\mathbf{z}|\boldsymbol{\theta}; \mathbf{x})$ and the reference distribution $\mathrm{p}(\mathbf{z}|\hat{\boldsymbol{\theta}}; \mathbf{x})$.

3 The exact integrated complete-data likelihood

Defined in a Bayesian perspective, the integrated complete-data likelihood of a mixture is defined by

$$\mathrm{p}(\mathbf{x}, \mathbf{z}) = \int_\Theta \mathrm{p}(\mathbf{x}, \mathbf{z}; \boldsymbol{\theta}) \mathrm{p}(\boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{10}$$

$\mathrm{p}(\boldsymbol{\theta})$ being the prior distribution of the model parameter $\boldsymbol{\theta}$. For much mixture models, this quantity is difficult to calculate and a BIC-like asymptotic

approximation can be used. It is given by

$$\ln \mathrm{p}(\mathbf{x},\mathbf{z}) = \ln \mathrm{p}(\mathbf{x},\mathbf{z};\hat{\boldsymbol{\theta}}) - \frac{\nu}{2}\ln n + O_{\mathrm{p}}(1) \tag{11}$$

where ν is the number of parameters to be estimated and $\hat{\boldsymbol{\theta}}$ corresponds to the ml of $\boldsymbol{\theta}$ obtained from the observed data $\mathbf{x}$ (since $\hat{\boldsymbol{\theta}}$ and the ml estimate of $\boldsymbol{\theta}$ obtained from the complete data $(\mathbf{x},\mathbf{z})$ are both consistent). Replacing the missing cluster indicators $\mathbf{z}$ by their Maximum A Posteriori (MAP) values $\hat{\mathbf{z}}$ for $\hat{\boldsymbol{\theta}}$ defined by

$$\hat{z}_{ik} = \begin{cases} 1 & \text{if } \arg\max_\ell t_{i\ell}(\hat{\boldsymbol{\theta}}) = k \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

Biernacki et al. (2000) obtained the following ICLbic criterion

$$\mathrm{ICLbic} = \ln \mathrm{p}(\mathbf{x},\hat{\mathbf{z}};\hat{\boldsymbol{\theta}}) - \frac{\nu}{2}\ln n. \tag{13}$$

This criterion aims favoring mixture situations giving rise to a partitioning of the data with the greatest evidence and, as a consequence, it appears to be robust against model misspecification (see Biernacki et al. 2000, and the experiments in the present paper).

Fortunately, in the context of multivariate multinomial distributions, there is no need to use such an asymptotic approximation because conjugate Jeffreys non informative prior distributions for all the parameters are available. Thus, the integrated complete-data likelihood (10) is closed form as shown hereunder.

Jeffreys non informative Dirichlet prior distributions for the mixing proportions and the latent class parameters are

$$\mathrm{p}(\boldsymbol{\pi}) = \mathcal{D}_g(\tfrac{1}{2},\ldots,\tfrac{1}{2}) \quad \text{and} \quad \mathrm{p}(\boldsymbol{\alpha}_k^j) = \mathcal{D}_{m_j}(\tfrac{1}{2},\ldots,\tfrac{1}{2}). \tag{14}$$

Assuming independence between prior distributions of the mixing proportions $\boldsymbol{\pi}$ and the latent class parameters $\boldsymbol{\alpha}_k^j$ $(k=1,\ldots,g;\ j=1,\ldots,d)$, it is straightforwardly get, using the conjugate property of the Multinomial-Dirichlet distributions (see for instance Robert 2001), that

$$\mathrm{p}(\mathbf{x},\mathbf{z}) = \frac{\Gamma(\frac{g}{2})}{\Gamma(\frac{1}{2})^g}\frac{\prod_{k=1}^g \Gamma(n_k+\frac{1}{2})}{\Gamma(n+\frac{g}{2})}\prod_{k=1}^g\prod_{j=1}^d \frac{\Gamma(\frac{m_j}{2})}{\Gamma(\frac{1}{2})^{m_j}}\frac{\prod_{h=1}^{m_j}\Gamma\left(n_k^{jh}+\frac{1}{2}\right)}{\Gamma(n_k+\frac{m_j}{2})}. \tag{15}$$

Replacing the missing labels $\mathbf{z}$ by $\hat{\mathbf{z}}$ in $\ln \mathrm{p}(\mathbf{x},\mathbf{z})$, as done to define the ICLbic criterion, the so-called ICL criterion is defined as follows:

$$\begin{aligned} \mathrm{ICL} &= \ln \mathrm{p}(\mathbf{x},\hat{\mathbf{z}}) \\ &= \sum_{k=1}^g \ln\Gamma(\hat{n}_k+\tfrac{1}{2}) + \sum_{k=1}^g\sum_{j=1}^d\left\{\sum_{h=1}^{m_j}\ln\Gamma\left(\hat{n}_k^{jh}+\tfrac{1}{2}\right) - \ln\Gamma(\hat{n}_k+\tfrac{m_j}{2})\right\} \\ &\quad + \ln\Gamma(\tfrac{g}{2}) - g\ln\Gamma(\tfrac{1}{2}) - \ln\Gamma(n+\tfrac{g}{2}) \\ &\quad + g\sum_{j=1}^d\left\{\ln\Gamma(\tfrac{m_j}{2}) - m_j\ln\Gamma(\tfrac{1}{2})\right\}, \end{aligned} \tag{16}$$

where $\hat{n}_k = \#\{i : \hat{z}_{ik}=1\}$ and $\hat{n}_k^{jh} = \#\{i : \hat{z}_{ik}=1, x_i^{jh}=1\}$.

4 The approximate integrated observed-data likelihood

The integrated observed-data likelihood (or integrated likelihood in short) is

$$\mathrm{p}(\mathbf{x}) = \int_{\Theta} \mathrm{p}(\mathbf{x};\boldsymbol{\theta})\mathrm{p}(\boldsymbol{\theta})d\boldsymbol{\theta}, \tag{17}$$

and a standard asymptotic approximation is given by

$$\ln \mathrm{p}(\mathbf{x}) = \ln \mathrm{p}(\mathbf{x};\hat{\boldsymbol{\theta}}) - \frac{\nu}{2}\ln n + O_{\mathrm{p}}(1), \tag{18}$$

which leads to the BIC criterion (Schwarz 1978)

$$\mathrm{BIC} = \ln \mathrm{p}(\mathbf{x};\hat{\boldsymbol{\theta}}) - \frac{\nu}{2}\ln n. \tag{19}$$

An approximate computation by importance sampling Denotipppppppng by $\mathcal{Z}$ all possible combinations of labels $\mathbf{z}$, Equation (17) can be written

$$\mathrm{p}(\mathbf{x}) = \sum_{\mathbf{z}\in\mathcal{Z}} \mathrm{p}(\mathbf{x},\mathbf{z}). \tag{20}$$

Since the integrated complete-data likelihood $\mathrm{p}(\mathbf{x},\mathbf{z})$ can be exactly calculated for the latent class model (see the previous section), the integrated likelihood $\mathrm{p}(\mathbf{x})$ is explicit.

Unfortunately, the sum over $\mathcal{Z}$ includes generally two many terms to be exactly computed. Following Casella et al. (2000), an importance sampling procedure can solve this problem. The importance sampling function, denoted by $I_{\mathbf{x}}(\mathbf{z})$, is a pdf on $\mathbf{z}$ ($\sum_{\mathbf{z}\in\mathcal{Z}} I_{\mathbf{x}}(\mathbf{z}) = 1$ and $I_{\mathbf{x}}(\mathbf{z}) \geq 0$) which can depend on $\mathbf{x}$, its support necessarily including the support of $\mathrm{p}(\mathbf{x},\mathbf{z})$. Denoting by $\mathbf{z}^1,\ldots,\mathbf{z}^s$ an i.i.d. sample from $I_{\mathbf{x}}(\mathbf{z})$, $\mathrm{p}(\mathbf{x})$ can be consistently estimated by the following Monte Carlo approximation

$$\hat{\mathrm{p}}(\mathbf{x}) = \frac{1}{S}\sum_{s=1}^{S} \frac{\mathrm{p}(\mathbf{x},\mathbf{z}^s)}{I_{\mathbf{x}}(\mathbf{z}^s)}. \tag{21}$$

This estimate is unbiased and its variation coefficient is given by

$$\mathrm{c}_v[\hat{\mathrm{p}}(\mathbf{x})] = \frac{\sqrt{\mathrm{Var}[\hat{\mathrm{p}}(\mathbf{x})]}}{\mathrm{E}[\hat{\mathrm{p}}(\mathbf{x})]} = \sqrt{\frac{1}{S}\left(\sum_{\mathbf{z}\in\mathcal{Z}} \frac{\mathrm{p}^2(\mathbf{z}|\mathbf{x})}{I_{\mathbf{x}}(\mathbf{z})} - 1\right)}. \tag{22}$$

In order to approximate the ideal importance function $I_{\mathbf{x}}(\mathbf{z}) = \mathrm{p}(\mathbf{z}|\mathbf{x})$ (*i.e.* the one which minimises the variance), two strategies are proposed.

- A maximum likelihood strategy consists of choosing the following estimate

$$\hat{\mathrm{p}}(\mathbf{z}|\mathbf{x}) = \mathrm{p}(\mathbf{z}|\mathbf{x};\hat{\boldsymbol{\theta}}) = \prod_{i,k}\left(t_{ik}(\hat{\boldsymbol{\theta}})\right)^{z_{ik}}. \tag{23}$$

- A Bayesian strategy consists of estimating more precisely

$$\mathrm{p}(\mathbf{z}|\mathbf{x}) = \int_{\Theta} \mathrm{p}(\mathbf{z}|\mathbf{x};\boldsymbol{\theta})\mathrm{p}(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta}$$

  by a Monte Carlo integration

$$\hat{\mathrm{p}}(\mathbf{z}|\mathbf{x}) = \frac{1}{R}\sum_{r=1}^{R} \mathrm{p}(\mathbf{z}|\mathbf{x};\boldsymbol{\theta}^r), \tag{24}$$

  where $\{\boldsymbol{\theta}^r\}$ are independent realisations of $\mathrm{p}(\boldsymbol{\theta}|\mathbf{x})$. In practice, a Gibbs sampler including the relabelling procedure described at the end of Section 2 is used.

These two strategies lead to two different criteria respectively called ILml and ILbayes (IL for Integrated Likelihood). The first one is parameterized by S whereas the second one depends on both S and R.

Remark As previously noticed, the support of the importance sampling function $\hat{\mathrm{p}}(\mathbf{z}|\mathbf{x})$ needs to include the support of $\mathrm{p}(\mathbf{z}|\mathbf{x})$ in order to avoid infinite variance in (22). For sufficiently large R, the Bayesian strategy ensures this property. Although it seems difficult to conclude about the maximum likelihood strategy, numerous experiments presented later suggest a good practical behaviour for this strategy.

Link between ICL and the integrated likelihood The following straightforward relationship exists between the integrated complete-data and observed-data likelihoods:

$$\ln \mathrm{p}(\mathbf{x},\hat{\mathbf{z}}) = \ln \mathrm{p}(\mathbf{x}) + \ln \mathrm{p}(\hat{\mathbf{z}}|\mathbf{x}). \tag{25}$$

Thus, as already noticed in Biernacki et al. (2000), the ICL criterion defined in (16) can be interpreted as the integrated likelihood penalized by a measure of the cluster overlap. It means that ICL tends to realize a compromise between the adequacy of the model to the data measured by $\ln \mathrm{p}(\mathbf{x})$ and the evidence of data partitioning measured by $\ln \mathrm{p}(\hat{\mathbf{z}}|\mathbf{x})$. For instance, highly overlapping mixture components leads typically to a low value of $\mathrm{p}(\hat{\mathbf{z}}|\mathbf{x})$ and consequently does not favor a high value of ICL.

In addition, the penalization $\mathrm{p}(\hat{\mathbf{z}}|\mathbf{x})$ can be regarded as the posterior mean of a particular utility function. As a matter of fact, it can be written

$$\mathrm{p}(\hat{\mathbf{z}}|\mathbf{x}) = \int_{\Theta} U(\hat{\mathbf{z}},\boldsymbol{\theta})\mathrm{p}(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta} \quad \text{with} \quad U(\hat{\mathbf{z}},\boldsymbol{\theta}) = \mathrm{p}(\hat{\mathbf{z}}|\mathbf{x};\boldsymbol{\theta}) = \prod_{i,k}(t_{ik}(\boldsymbol{\theta})^{\hat{z}_{ik}}, \tag{26}$$

where $U(\cdot,\cdot)$ is the utility function. At this point, a remark is to be made. From a decision-theoretic point of view, a utility function lies on the space $\mathcal{D}\times\Theta$ where $\mathcal{D}$ corresponds to the so-called *decision* space. Generally, $\mathcal{D} = \Theta$ and the utility function can be seen as a measure of proximity between its two arguments in Θ. Here, the decision space corresponds to a classification $\mathbf{z}$ of the observed data $\mathbf{x}$. It is related to the parameter space Θ but in an indirect way since $\hat{\mathbf{z}} = \mathrm{MAP}(\hat{\boldsymbol{\theta}})$.

5 Numerical experiments

We illustrate the behaviour of the non asymptotic criteria for simulated data for which we distinguish two different situations: A situation where the data arose from one of the mixtures in competition and a situation where the latent class model did not give rise to the data. Finally, we treat an example on a real data set.

5.1 Simulated data: Well specified model

Design of experiments Observations are described by six variables ($d = 6$) with numbers of modalities $m_1 = \ldots = m_4 = 3$ and $m_5 = m_6 = 4$. Two different numbers of mixture components are considered: A two component mixture ($g = 2$) with unbalanced mixing proportions, $\boldsymbol{\pi} = (0.3\ 0.7)'$, and a four component mixture ($g = 4$) with equal mixing proportions, $\boldsymbol{\pi} = (0.25\ 0.25\ 0.25\ 0.25)'$. In each situation, three values of the parameter $\boldsymbol{\alpha}$ are chosen to get a low, a moderate and a high cluster overlapping, respectively defined as 15%, 30% and 60% of the worst possible error rate (situation where $\boldsymbol{\alpha}_k^{jh} = 1/m_j$). For the previous structures associated to $g = 2$ and $g = 4$, this worst error rate is 0.30 and 0.75 respectively. More precisely, the chosen structure for $\boldsymbol{\alpha}$ is expressed by

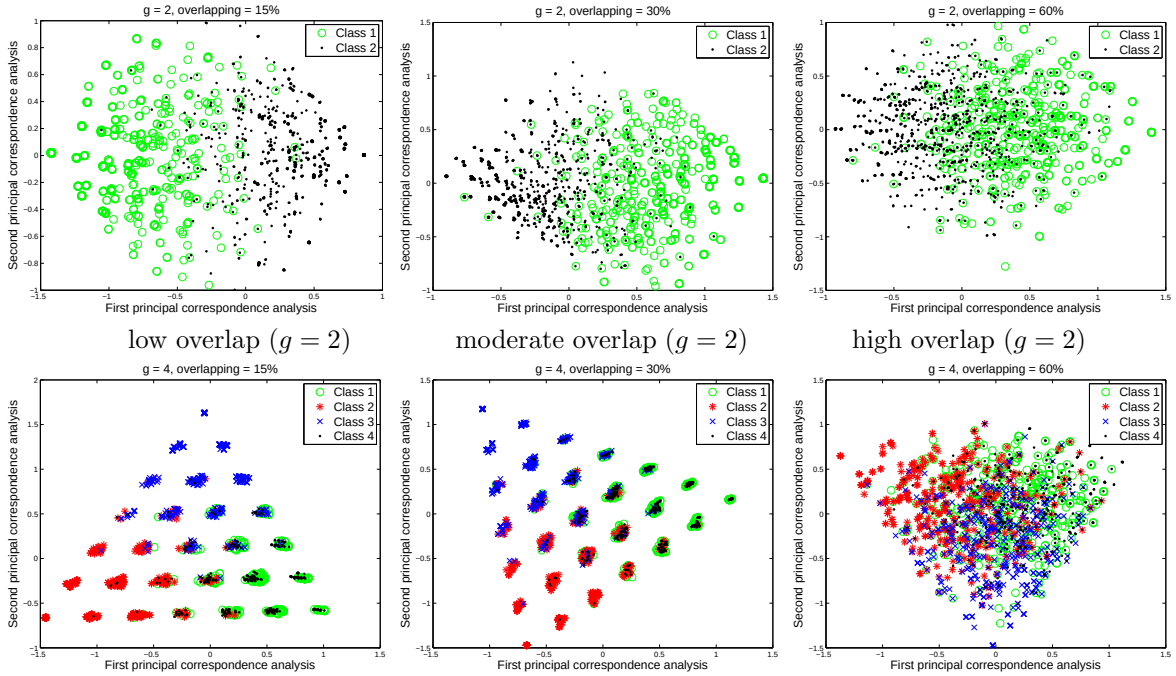
$$\alpha_k^{jh} = \begin{cases} \frac{1}{m_j} + (1-\delta)\frac{m_j-1}{m_j} & \text{if } h = \Big[(k-1) \text{ modulo } m_j\Big] + 1 \\ \frac{\left(1-\frac{1}{m_j}-(1-\delta)\frac{m_j-1}{m_j}\right)}{m_j-1} & \text{otherwise,} \end{cases} \tag{27}$$

where $0 \leq \delta \leq 1$ allows to fit mixture parameters with the required overlapping: $\delta = 0$ corresponds to the minimum overlap because $\alpha_k^{jh} = 0$ or 1, whereas $\delta = 1$ corresponds to the maximum overlap because $\alpha_k^{jh} = 1/m_j$. Since the overlap is a continuous and non decreasing function of δ, the value $\boldsymbol{\alpha}$ associated to a given overlap is easily derived from a numerical procedure. Table 1 provides computed values of δ for each situation. In addition, Figure 1 displays a data sample for $g = 2$ and $g = 4$ on the first two axes of a correspondence analysis.

		$g = 2$		$g = 4$	
overlap	% of max.	error rate	δ	error rate	δ
low	15	0.0450	0.4713	0.1125	0.4770
moderate	30	0.0900	0.5822	0.2250	0.6097
high	60	0.1800	0.7313	0.4500	0.7900
maximum	100	0.3000	1.0000	0.7500	1.0000

Table 1: Error rate and corresponding value of δ for each parameter structure. The reference overlap case (denoted by "maximum"), corresponding to the worst possible error rate, is also given.

Results for the ICL criteria For each parameter structure, 20 samples are generated for four different sample sizes $n \in \{320, 1600, 3200, 16000\}$. For each sample and for a number of mixture components varying from $g = 1$ to



low overlap ($g = 2$) moderate overlap ($g = 2$) high overlap ($g = 2$)

low overlap ($g = 4$) moderate overlap ($g = 4$) high overlap ($g = 4$)

Figure 1: A sample ($n = 1600$) arising from each $g \in \{2, 4\}$ mixture situation for low, moderate and high overlapping. It is displayed on the first plane of a correspondence analysis and an i.i.d. uniform noise on $[0, 0.01]$ has been added on both axes for each point in order to clarify the visualization.

6, the EM algorithm has been run 11 times with random initial parameters (uniform distribution on the parameter space) for a sequence of 1000 iterations and the best run is retained as being the maximum likelihood estimate. Relative frequency of choosing the number of mixture components with criteria ICL and ICLbic is displayed on Figures 2 and 3 respectively for $g = 2$ and $g = 4$. In addition, the BIC criterion is provided on the same figures. ICL-type and IL-type criteria are compared in Subsection 5.2.

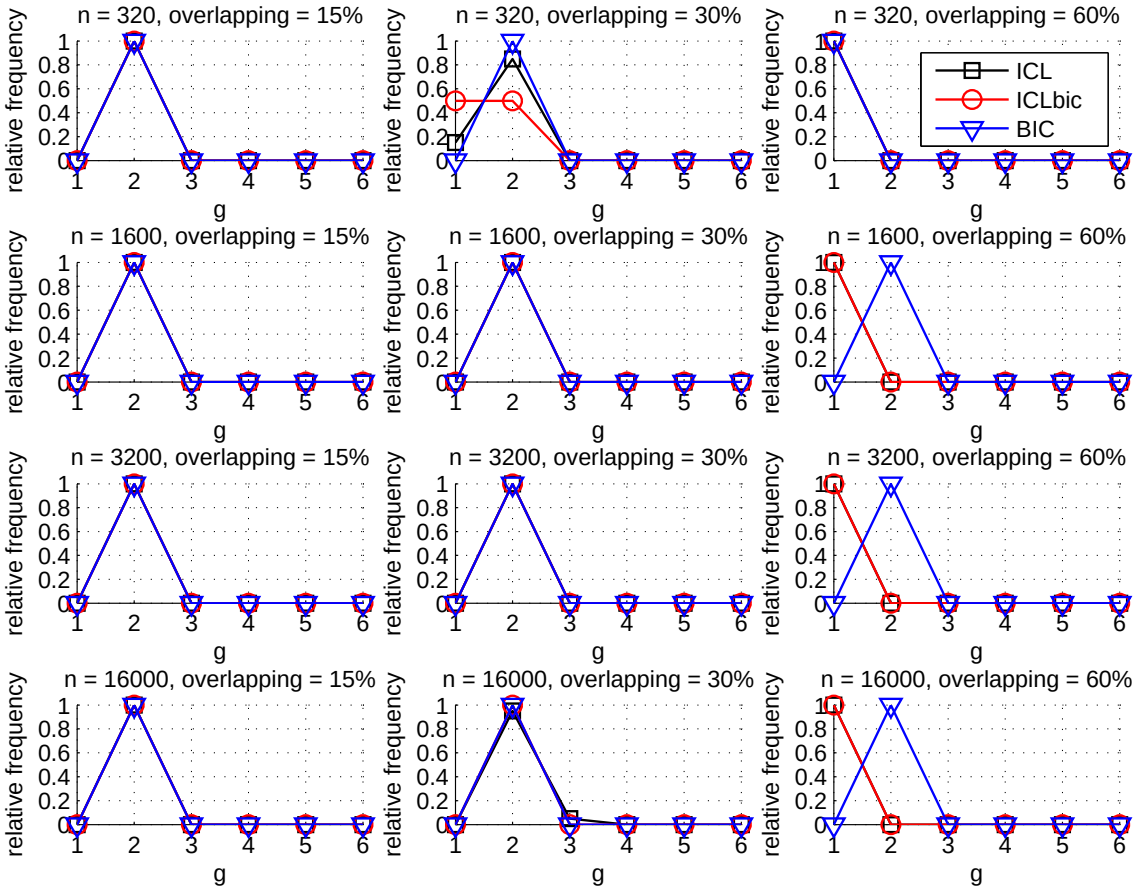


Figure 2: Relative frequency of choosing the number of mixture components with ICL-type criteria when $g = 2$.

As expected, it appears that ICL and ICLbic behave the same for large sample sizes. Sometimes, asymptotic behaviour of both criteria is reached for small sample sizes (low and high overlap situations). However, when asymptotic behaviour is reached only for larger sample sizes (typically for moderate overlap situations), ICL converges far faster than ICLbic towards its limit. We also notice than, before reaching its asymptotic behaviour, ICLbic is much more conservative than ICL since it detects less components than ICL. Thus, ICL can be preferred to ICLbic since it behaves better and is not really more complex to compute.

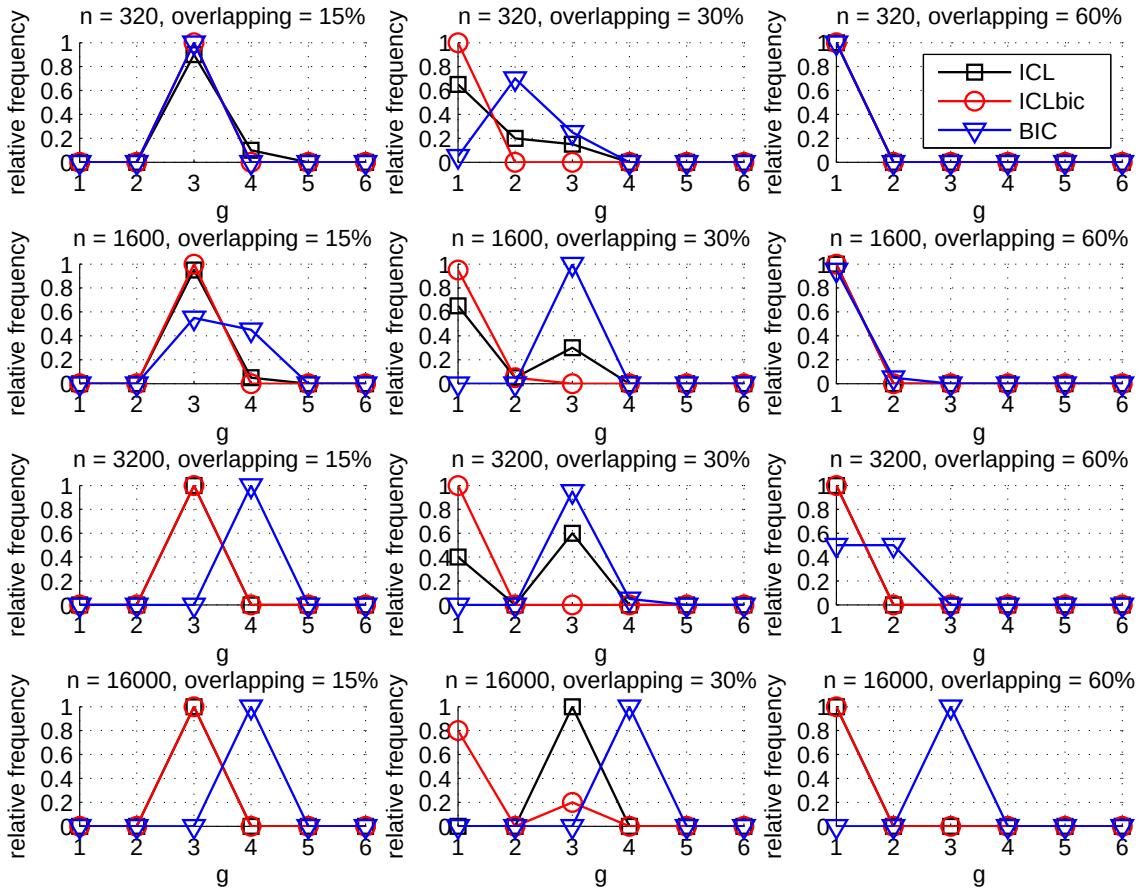


Figure 3: Relative frequency of choosing the number of mixture components with ICL-type criteria when $g = 4$.

Results for the IL criteria From Figure 2, it is clear that BIC reached its asymptotic behaviour for small sample sizes in the $g = 2$ components situation. Thus, it is more informative to focus on situations involving $g = 4$ components. The same samples and experimental conditions than previously defined are used. In addition the Gibbs sampler, initialised at random from a uniform distribution on the parameter space, generates a sequence of 11,000 parameters, the first 1000 draws corresponding to the *burn-in* period. The R values $\boldsymbol{\theta}^r$ are selected in the remaining sequence of size 10,000 every $1000/R$ draws. Since values $R = 50$ and $R = 100$ are retained, it guarantees that the selected draws are quasi independent. A value of $\boldsymbol{\theta}^r$ is selected in the remaining sequence of size 10,000 every 100 draws when $R = 100$, and every 200 draws when $R = 50$.

Figure 4 displays relative frequency of choosing the number of mixture components for all IL-type criteria. It includes the asymptotic criterion BIC, the non asymptotic criteria ILml and ILbayes, and also a naive criterion called ILu with makes use of a uniform importance sampling function, $I_{\mathbf{x}}(\mathbf{z}) = 1/\#\mathcal{Z}$ for estimating $\mathrm{p}(\mathbf{x})$. In each case, $S = 1000$ and, moreover, for ILbayes, $R = 100$.

Selected values for S and R appear to be sufficiently large for allowing BIC, ILml and ILbayes to behave the same for large n. However, the naive ILu criterion is clearly disqualified since it always selects one component (a much greater value of S is certainly required). It highlights the importance of choosing a sensible importance sampling function. Moreover, as for ICL comparisons, ILml and ILbayes criteria reach their asymptotic behaviour far faster than BIC. Thus, it illustrates again the interest of non asymptotic approximations of the integrated likelihood for the latent class model. Finally, ILbayes outperforms ILml in experiments involving a small sample size ($n = 320$), but both criteria quickly behave the same when n increases.

Figure 5 evaluates the influence of R and S on the ILbayes behaviour. It appears that variability of the criterion is not really significant for R and S values in this range. Similarly, Figure 6 estimates the influence of S on both ILml and ILu performances. Again, no significant variability can be identified in the considered range for S.

5.2 Simulated data: Misspecified model

In this subsection, we focused on the difference which could occur in practice between IL and ICL criteria.

From Figures 2 and 3, it is apparent that ICL criteria have a tendency to underestimate the right number of components. This tendency appears more clearly in the high overlap case where ICL always underestimates the right number of clusters, even for large n. In this case, the entropic penalty term in ICL is high and actually there is no evidence for data partitioning (see the right column in Figure 1). The realistic case where the data are not following a latent class model is now considered. It will allow us to highlight the possible interest of ICL in a cluster analysis context.

Design of experiments Two well separated components (about 0.07 error rate) are considered in a situation where the conditional independence assumption is not true. Data have been generated with the following procedure:

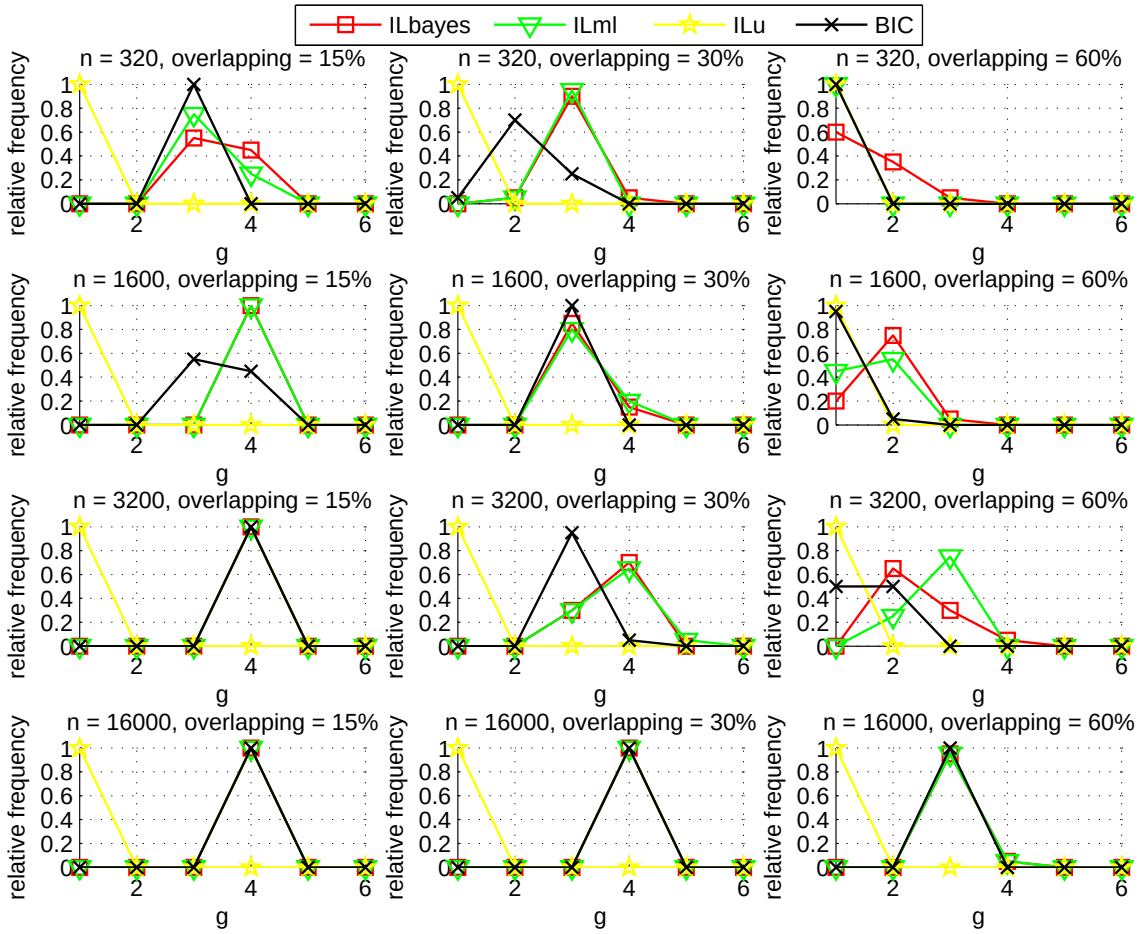


Figure 4: Relative frequency of choosing the number of mixture components when $g = 4$ for all IL criteria with $S = 1000$: BIC, ILu, ILml and ILbayes ($R = 100$ for this latter).

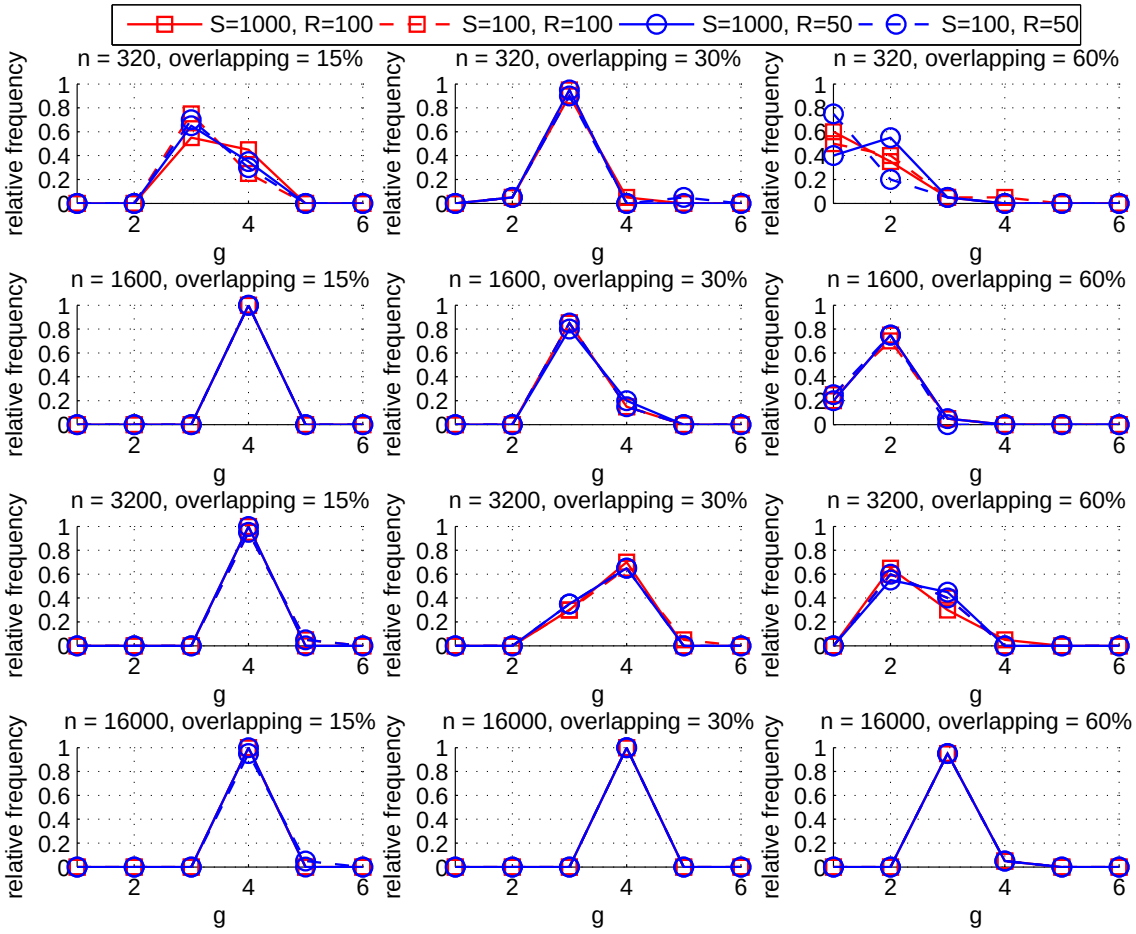


Figure 5: Relative frequency of choosing the number of mixture components when $g = 4$ for ILbayes criterion with different values for R and S.

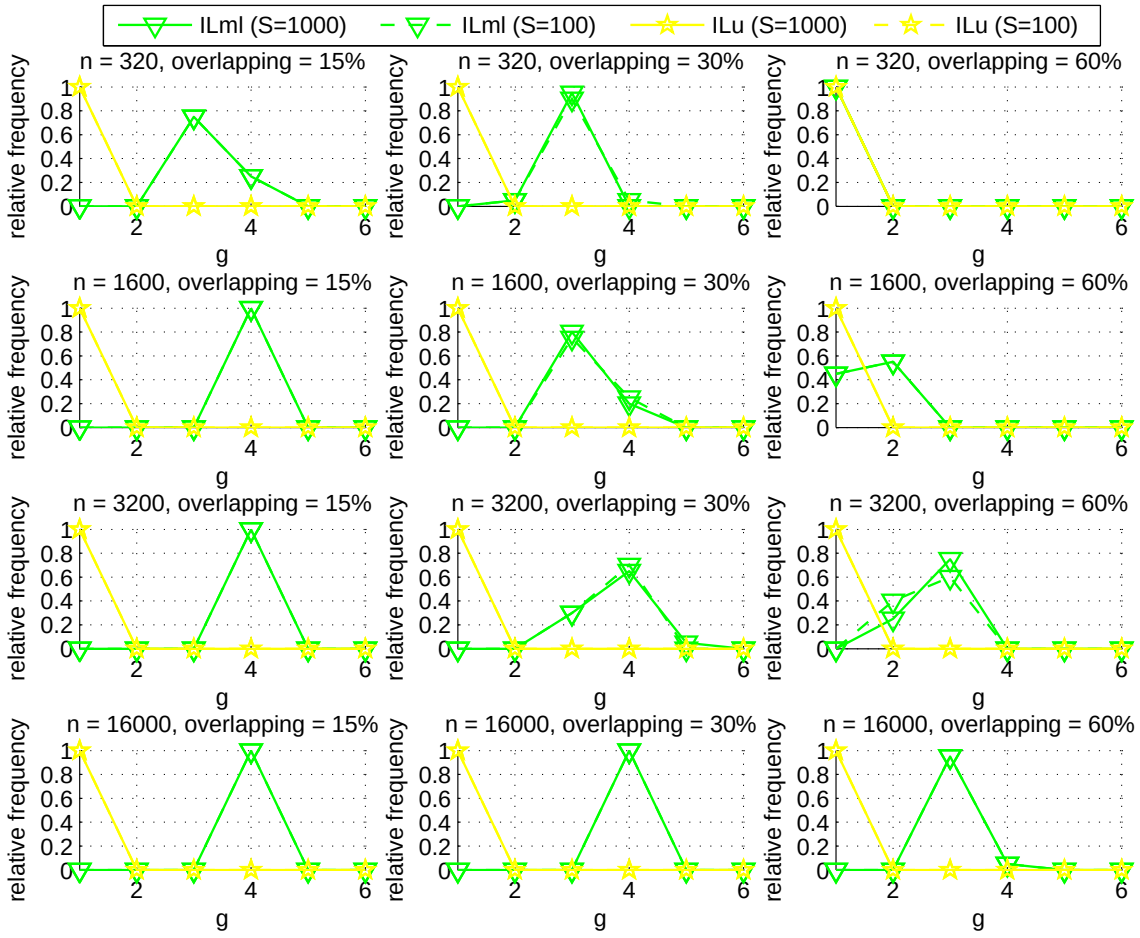


Figure 6: Relative frequency of choosing the number of mixture components when $g = 4$ for ILml and ILu criteria with different values for S.

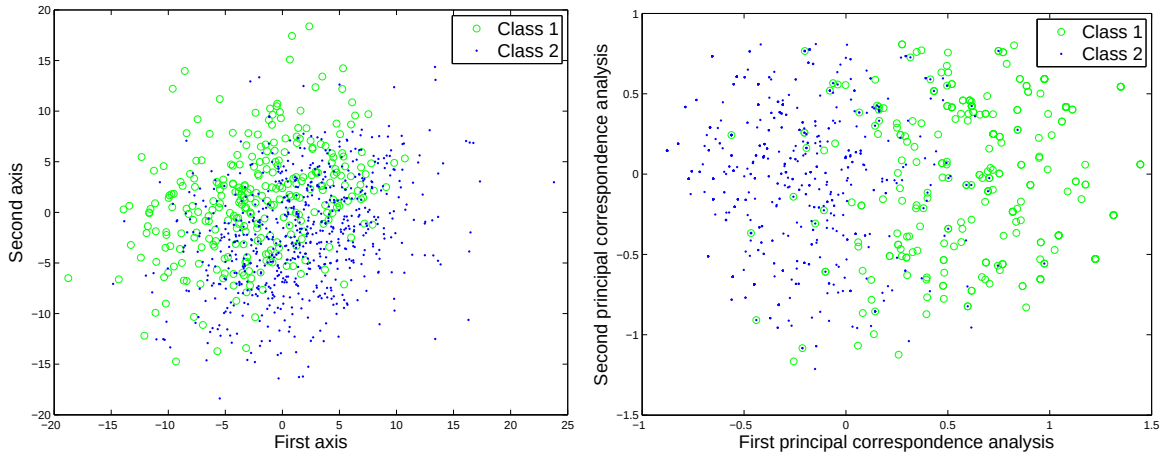
1. Firstly, a sample of size n is drawn from a two component Gaussian mixture in $\mathbb{R}^6$ with mixing proportions $\boldsymbol{\pi} = (0.3\ 0.7)'$, with centers $\boldsymbol{\mu}_1 = (-2\ 2\ -2\ -2\ -2\ -2)'$ and $\boldsymbol{\mu}_2 = (2\ -2\ 2\ 2\ 2\ 2)'$ and with variance matrices $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \mathbf{DAD}'$ where

$$\mathbf{A} = 10 \times \begin{bmatrix} 4 & 0 & \mathbf{0}_4' \\ 0 & 2 & \mathbf{0}_4' \\ \mathbf{0}_4 & \mathbf{0}_4 & \mathbf{I}_4 \end{bmatrix} \text{ and } \mathbf{D} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & \mathbf{0}_4' \\ 1/\sqrt{2} & -1/\sqrt{2} & \mathbf{0}_4' \\ \mathbf{0}_4 & \mathbf{0}_4 & \mathbf{I}_4 \end{bmatrix}. \quad (28)$$

   The four-variate identity matrix is denoted by $\mathbf{I}_4$ and $\mathbf{0}_4$ denotes the four-variate zero vector. It is to be noticed that conditional independence between axes 1 and 2 is broken since they are correlated for both mixture components.

2. Then, $\mathbb{R}^6$ is discretized in the following manner in order to obtain categorical data: (1) axes 1 to 4 are divided into three modalities $]-\infty, -2[$, $[-2, 2[$ and $[2, \infty[$, (2) axes 5 and 6 are divided into four modalities $]-\infty, -1[$, $[-1, 0[$, $[0, 1[$ and $[1, \infty[$. Thus, the same dimension space and number of modalities per variable that in the simulated data of Section 5.1 is retrieved.

Figure 7 displays a sample before and after discretization. The other experimental conditions are similar to those ones considered in Section 5.1, excepted that five different sample sizes are retained ($n \in \{320, 1600, 3200, 16000, 80000\}$) and that 30 samples are generated instead of 20 for each situation.



(a) (b)

Figure 7: A sample from the correlated situation (a) *before* discretization on the first two canonical axes and (b) *after* discretization on the first two correspondence analysis axes.

Results Frequency of choosing different g values is displayed on Figure 8 for BIC, ICL and ICLbic. It clearly appears that ICL and ICLbic favor two groups for any sample size whereas BIC prefers a higher number of components when the sample size significantly increases. It illustrates the robustness of

ICL criteria already noticed in the Gaussian situation by Biernacki et al. (2000) where ICLbic was able to select well separated clusters even when the model was misspecified. On the contrary, the BIC criterion is focused on detecting latent classes providing a good fit of the mixture with the data without considering the cluster overlap.

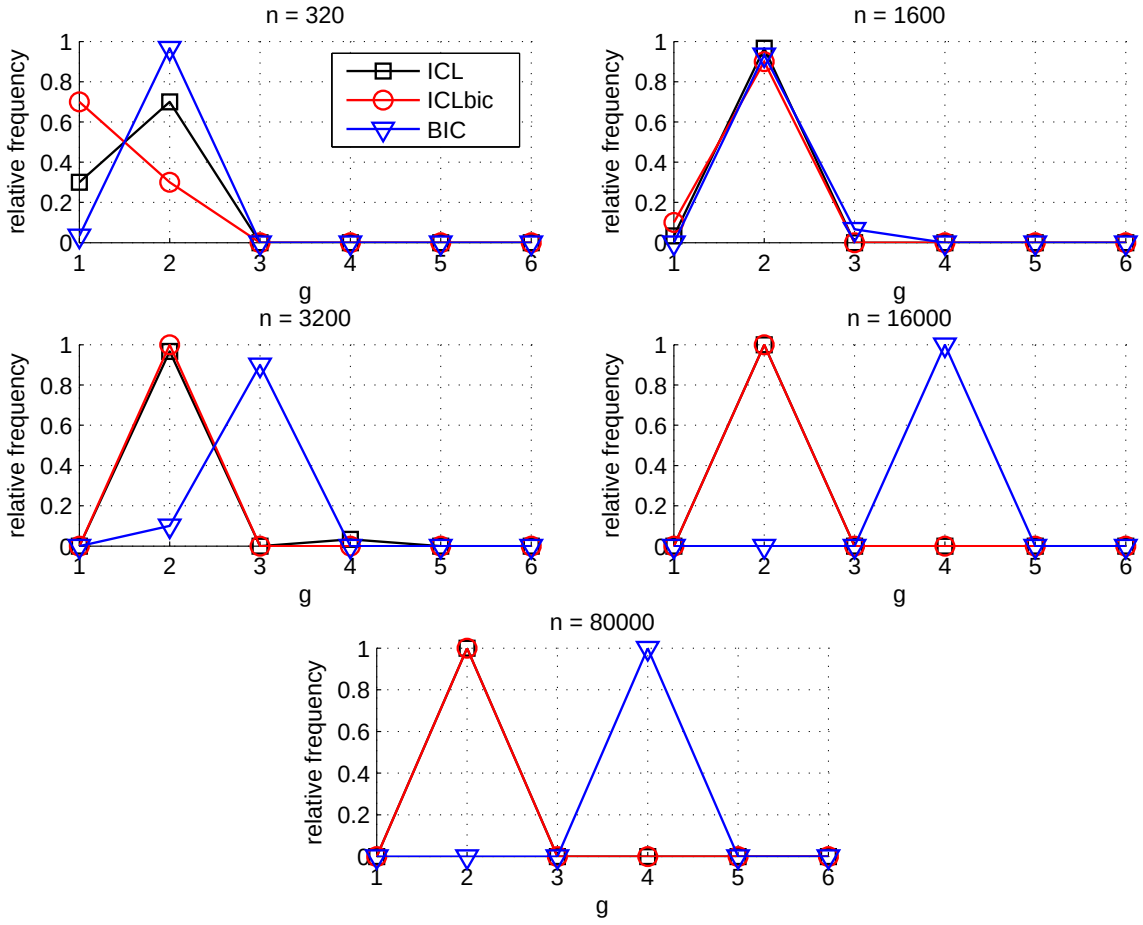


Figure 8: Relative frequency of choosing the number of groups when the conditional independence is not verified.

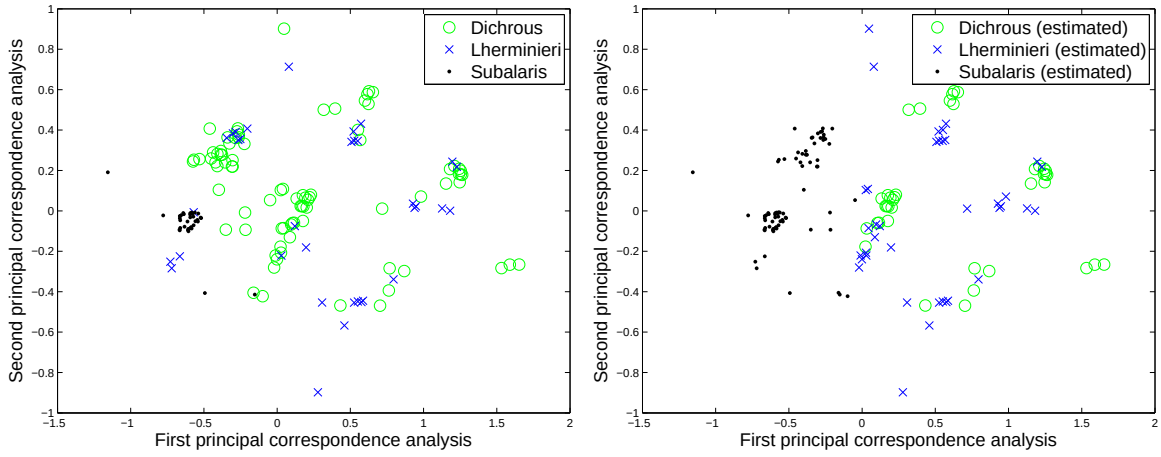
5.3 A biological data set

The data Puffins are pelagic seabirds from the family Procellaridae. A data set of 153 puffins divided into three subspecies *dichrous* (84 birds), *lherminieri* (34 birds) and *subalaris* (35 birds) is considered (Bretagnolle 2007). These birds are described by the five plumage and external morphological characters displayed in Table 2. Figure 9 (a) displays the birds on the first correspondence analysis plan.

	modalities				
variables	1	2	3	4	5
gender	male	female			
eyebrows⋆	none	.....		very pronounced	
collar⋆	none	.....			continuous
sub-caudal	white	black	black & white	black & WHITE	BLACK & white
border⋆	none	.....	many		

⋆ using a paper pattern

Table 2: Details of plumage and external morphological characters for the seabird data set.



(a) (b)

Figure 9: Seabirds on the first two correspondence analysis axes (a) with the *true partition* and (b) with the *EM estimated partition*. An i.i.d. uniform noise on $[0, 0.1]$ has be added on both axes for each individual in order to improve visualization.

Results for ICL criteria For number of groups varying from $g = 1$ to 6, EM is run 10 times at random (uniform distribution on the parameter space) for 1000 iterations and the run providing the largest likelihood is considered as the ml estimate. Table 3 displays values of ICL, ICLbic and BIC for each number of components. It appears that only ICL selects three groups. The corresponding estimated partition, where labels are chosen to ensure the minimum error rate with the true partition, is given in Figure 9 (b) and described also in Table 4. It has to be compared with the true partition given in Figure 9 (a). It leads to 55 misclassified birds (35.95% of birds), a Rand criterion value of 0.6121 and a corrected Rand criterion value of 0.1896 (Rand 1971).

	g					
criteria	1	2	3	4	5	6
ICL	-712.0771	-712.5665	**-711.8138**	-727.4411	-737.4558	-741.7878
ICLbic	**-714.0339**	-727.3274	-735.7768	-774.0148	-799.3678	-830.8298
BIC	-714.0339	**-711.1445**	-730.3857	-754.5809	-784.8988	-814.6092

Table 3: Value of ICL criteria and BIC for different number of groups on the seabird data set. Boldface indicates maximum value for each criterion.

	$\hat{\mathbf{z}}$		
$\mathbf{z}$	*dichrous*	*lherminieri*	*subalaris*
dichrous	39	14	31
lherminieri	0	24	10
subalaris	0	0	35

Table 4: Confusion table (in number of individuals) between the true partition $\mathbf{z}$ and the *three groups* partition $\hat{\mathbf{z}}$ estimated from the EM solution.

On an other hand, it has to be noticed than ICL hesitates between one, two or three clusters. It seems to point out that there is little difference between the birds, and that it could be doubtful to discriminate the sub-species with the available variables. Moreover, it appears that ICLbic and BIC do not behave the same since ICLbic has a marked preference for the one component solution (no clustering) while BIC clearly favors the two clusters solution (Table 5 gives the related confusion table).

	$\hat{\mathbf{z}}$	
$\mathbf{z}$	group 1	group 2
dichrous	36	48
lherminieri	12	22
subalaris	35	0

Table 5: Confusion table (in number of individuals) between the true partition $\mathbf{z}$ and the *two groups* partition $\hat{\mathbf{z}}$ estimated from the EM solution.

Results for IL criteria Experiments are now focused on criteria ILml, ILbayes and ILu. The implemented Gibbs sampler is the same than with the simulated data sets. For different values of R ($R \in \{50, 100\}$) and S ($S \in \{100, 1000, 10000\}$), each criterion is computed 100 times. Figure 10 displays the mean of each criterion values over the 100 runs. More precisely, it provides the mean of the exponential of each criterion in order to work on estimates of $\mathrm{p}(\mathbf{z}|\mathbf{x})$ instead of estimates of $\ln \mathrm{p}(\mathbf{z}|\mathbf{x})$. In addition, Figure 11 provides this mean separately for each criterion. Figure 12 provides the variation coefficient of the exponential of all the criteria.

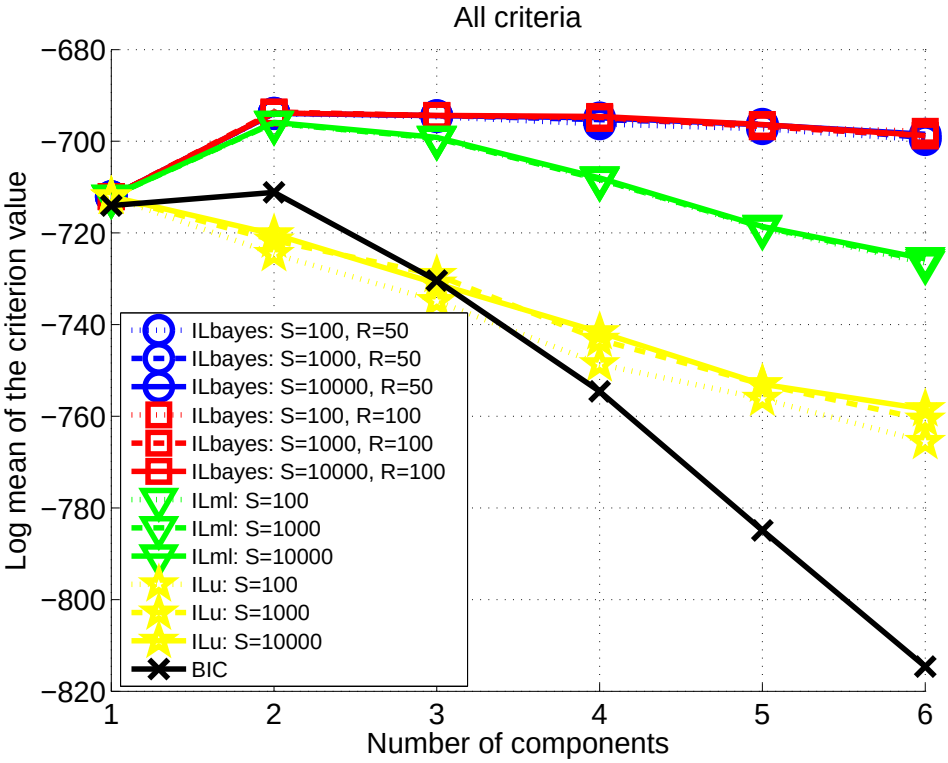


Figure 10: Representation of the mean of IL criteria for different R and S values.

Not surprisingly, the variability of all criteria tends to increase with the number of components. It is a consequence of the fact that the number of partitions $\mathcal{Z}$ increases with g. It could be meaningful to increase S with g.

It appears that exp(ILu) is the worst estimate of $\mathrm{p}(\mathbf{z}|\mathbf{x})$ since it has the largest variation coefficient. A very high value of S is certainly needed to reduce dramatically this variability. Here, ILu selects only one group.

As expected, the criterion with the smallest variation coefficient is ILbayes when S are R are large ($S = 10000$ and $R = 100$). In average, it selects two groups but criterion values for $g \in \{2, 3, 4\}$ are close. Smaller values of S or R do not significantly change the mean criterion but tend to increase its variability.

The ILml criterion could be seen as an interesting cheaper version of ILbayes. Although its variability is often larger than the ILbayes variability, it allows to

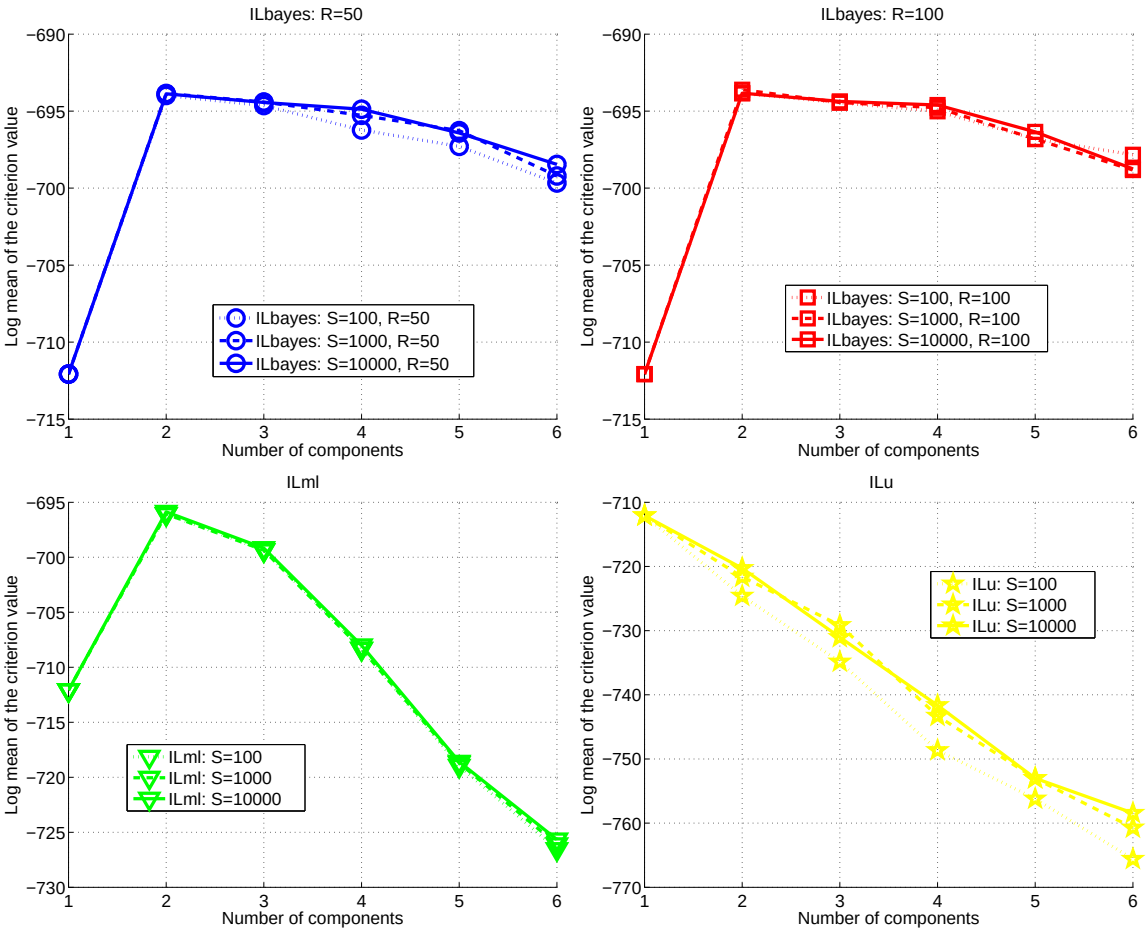


Figure 11: *Separate* representation of the mean of IL criteria for different R and S values.

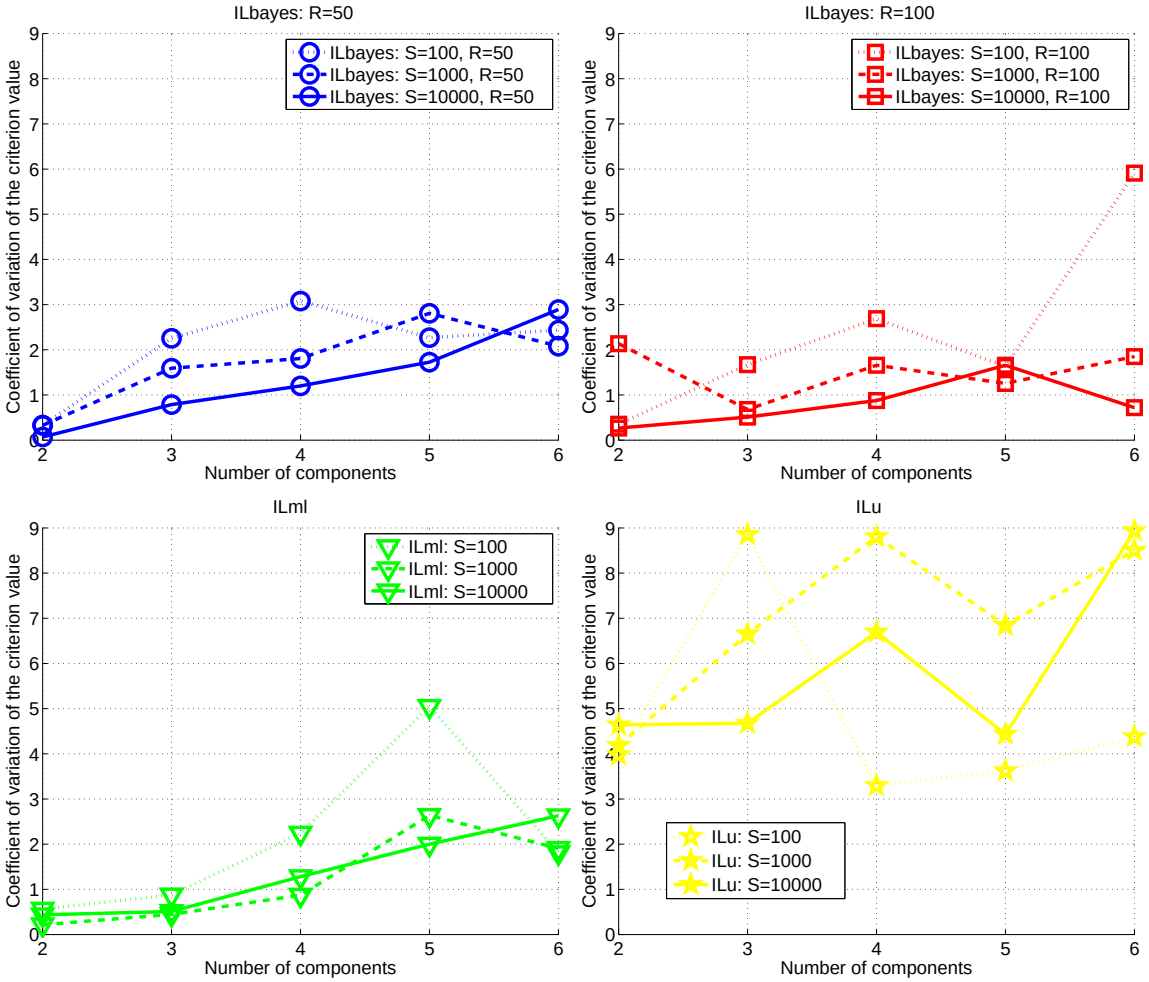


Figure 12: *Separate* representation of the variation coefficient of IL criteria for different R and S values (same scale is used for all sub-figures).

select also two clusters while preserving a lot of CPU time (Gibbs sampler and label-switching procedures are time consuming). Moreover, on the contrary to BIC, ILml does not exclude the possibility of choosing three groups.

6 Concluding remarks

In this paper, we exploit the fact that the Jeffreys non informative prior distribution of the parameters of the multivariate multinomial mixture model is a conjugate distribution. It implies that the integrated complete-data likelihood can be expressed explicitly. Moreover, it helps to derive a non asymptotic approximation of the integrated observed-data likelihood. Simple and efficient numerical procedures to get such a non asymptotic approximation are proposed.

Monte Carlo numerical experiments for selecting the number of groups in a latent class model highlight the interest of using exact or approximate non asymptotic criteria instead of standard asymptotic criteria as ICLbic or BIC. In particular, they illustrate the fact that asymptotic criteria may fail to detect interesting structures in the data for small sample sizes.

On another hand, this paper underlines the possible interest of using the integrated complete-data likelihood criterion rather than the integrated observed-data likelihood criterion. The first one explicitly favors models leading to well separated groups. This feature implies some robustness against model misspecification, as the violation of the conditional independence assumption of the latent class model.

From the encouraging results obtained for non asymptotic criteria in this latent class model context, it is now challenging to decline such criteria in other model-based situations. It includes for instance the possibility to design such criteria to variants on the latent class model considering constrained parameters to get more parsimonious models (see Celeux and Govaert 1991).

Acknowledgements Authors are indebted to the biologist Vincent Bretagnolle (CEBC-CNRS, Beauvoir sur Niort, France) for having provided the seabird data set. They are greteful to Jean-Michel Marin (Inria) for helpful discussions and remarks.

References

Aitkin, M., Anderson, D. and Hinde, J.: 1981, Statistical modelling of data on teaching styles (with discussion), *Journal of The Royal Statistical Society (series B)* **47**(1), 67–75.

Biernacki, C., Celeux, G. and Govaert, G.: 2000, Assessing a mixture model for clustering with the integrated completed likelihood, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(7), 719–725.

Bretagnolle, V.: 2007, Personal communication, source: Museum.

Casella, G., Robert, C. and Wells, M.: 2000, Mixture models, latent variables and partitioned importance sampling, *Technical Report 2000-03*, CREST, INSEE, Paris.

Celeux, G. and Govaert, G.: 1991, Clustering criteria for discrete data and latent class models, *Journal of Classification* **8**(2), 157–176.

Celeux, G., Hurn, M. and Robert, C. P.: 2000, Computational and inferential difficulties with mixture posterior distributions, *Journal of the American Statistical Association* **95**, 957–970.

Dempster, A. P., Laird, N. M. and Rubin, D. B.: 1977, Maximum likelihood from incomplete data via the EM algorithm (with discussion), *Journal of the Royal Statistical Society* **B 39**, 1–38.

Fraley, C. and Raftery, A. E.: 2002, Model-based clustering, discriminant analysis and density estimation, *Journal of the American Statistical Association* **97**, 611–631.

Goodman, L. A.: 1974, Exploratory latent structure analysis using both identifiable and unidentifiable models, *Biometrika* **61**, 215–231.

Marin, J.-M., Mergersen, K. and Robert, C. P.: 2005, *Bayesian modelling and inference on mixture of distributions*, Elsevier B. V., Handbbok of Statistics, Vol. 25.

McLachlan, G. J. and Krishnan, K.: 1997, *The EM Algorithm*, Wiley, New York.

McLachlan, G. J. and Peel, D.: 2000, *Finite Mixture Models*, Wiley, New York.

Nadif, M. and Govaert, G.: 1998, Clustering for binary data and mixture models: Choice of the model, *Applied Stochastic Models and Data Analysis* **13**, 269–278.

Rand, W. M.: 1971, Objective criteria for the evaluation of clustering methods, *Journal of American Statistical Association* **66**, 846–850.

Robert, C. P.: 2001, *The Bayesian Choice*, Springer Verlag, second edition, New York.

Schwarz, G.: 1978, Estimating the number of components in a finite mixture model, *Annals of Statistics* **6**, 461–464.

Stephens, M. A.: 2000, Dealing with label-switching in mixture models, *Journal of the Royal Statistical Society series B* **62**, 795–809.

Centre de recherche INRIA Sophia Antipolis – Méditerranée
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399